

LUD 5538.1 CIP - JEL/NDH (09807399)

Claim 139: The isolated nucleic acid molecule of claim 110, comprising SEQ ID NO: 8.

Claim 140: The isolated nucleic acid molecule of claim 110, consisting of SEQ ID NO: 7.

Claim 141: The isolated nucleic acid molecule of claim 110, consisting of SEQ ID NO: 6.

Claim 142: The isolated nucleic acid molecule of claim 110, consisting of SEQ ID NO: 7.

Claim 143: The isolated nucleic acid molecule of claim 110, consisting of SEQ ID NO: 8.

#### REMARKS

The preceding amendment is presented in accordance with 37 CFR § 1.121(h). Claims 110-143 will be pending.

The amendment address points 3 and 10 of the office action. The objections to claims 108 and 109 are obviated via the sequence listing. Applicants note that the sequence listing objection was not sent with the June 5 office action, but was telefaxed on June 12, 2001.

Prior to addressing the bases for rejecting the claims, it is worthwhile to discuss what the empirical evidence provided in the application shows.

In example 1, a melanoma cell line is discussed. The prior art describes how various members of the CT family of antigens are expressed by this cell. The cell was used to determine expression of other members of the family. It should be borne in mind that the term "CT antigen" is discussed in the specification.

The cells were studied, using the "SEREX" methodology. This method is known in the art. To perform SEREX, total RNA is removed from cells, converted to cDNA and, a cDNA library is constructed in a vector. In this case, "λZAP" was used. The λZAP cDNA library was then used to transform E. coli.

After E. coli are transformed, the transformed cells are used to screen samples for antibodies. Allogeneic samples were used in example 1. Essentially, one is determining if the

LUD 5538.1 CIP - JEL/NDH (09807399)

protein encoded by the cDNA used to transform the *E. coli* reacts with antibodies in the sample. One knows that the reaction is not with a "native" *E. coli* protein, because in the SEREX method, such antibodies are "stripped" prior to reaction with the transformants. Applicants refer the examiner to the references cited at page 5, i.e., *Sahin, et al, Proc. Natl. Acad Sci USA* 92:11810-11813, and U.S. Patent No. 5,698,396, both of which are incorporated by reference. Full details of the SEREX methodology are described in these references. Hence, one knows that, if there are any positives, the reactive protein was from the cDNA library.

The yield in this experiment was high, i.e., there were 65 positives.

Example 2 describes, simply, attempts to identify the positive materials. Standard methods were used.

Table 8, at page 8, shows that of the 65 positives 18 were known CT antigens. Hence, one knows from this that one can identify known members of the CT family. Ten sequences were not identifiable in public sequence libraries. There were also, 33 clones which belonged to the "KOC" family.

Examples 3 and 5 deal with CT antigen, i.e., "CT7." Please see US Patent No. 6,297,364.

Example 6 describes sequence analysis of the 33 members of the family found in the experiments of example 2. As this example makes manifestly clear, two different genes, referred to as KOC-2 and KOC-3 were identified. They were clearly not artifacts, because when a testicular cDNA library was analyzed, again using allogeneic samples, the molecules were found.

In example 7, the DNA expression pattern of the genes was studied. KOC-2 was found in normal tissue, but only testis and, with modification of the assay, in normal kidney cells. This is a "classic" CT expression pattern. As noted in the application, "CT" means "cancer/testis."

The KOC-3 gene was found to be expressed in both cancer and normal cells.

Now, the examiner has rejected all of claims 53-73, 80-84, 108 and 109, arguing that the claims are not supported by either "a specific substantial utility or a well-established utility."

The examiner states that "the specification fails to demonstrate a utility for the nucleic acids of SEQ ID NOS: 5-8, as the specification fails to correlated the presence of the protein encoded by SEQ ID NOS: 5-8 in clinical samples." The examiner then goes on to hypothesize that (i) antibodies can cross react with different proteins, and (ii) that the polynucleotides may

LUD 5538.1 CIP - JEL/NDH (09807399)

not be expressed due to any of a number of hypotheses. The examiner also states that "under conditions of recombinant expression, epitopes of KOC-2 and 3 which are not normally expressed in a patient within the context of MHC could be expressed by *E. coli*."

All of this is unproven speculation. The facts are the following. Something in the serum of patient NW38 reacted with a protein expressed by cDNA from an allogeneic source. Antibodies were present in the cancer patient's serum. They did not react with an *E. coli* protein because, as has been pointed out, supra, when SEREX is carried out, any reactivity with *E. coli* proteins is eliminated by use of controls. Again, the examiner is directed to the references which describe the SEREX methodology.

With respect to the elaborate argument regarding factors affecting translation of proteins, this is all well and good, but it is irrelevant. It is irrelevant because there were antibodies in the sample analyzed. Antibodies are produced in response to a protein. Hence, at some point, NW38 (the source of the sample), did have the protein in his or her blood. It got there by being expressed. The issue for CT antigens, as was set forth very clearly in the specification, is: are they expressed in individuals with cancer? The evidence indicates that KOC-2 and KOC-3 are so expressed, regardless of what happened to the irrelevant ferritin polypeptide, ornithine decarboxylase, and p glycoprotein discussed by the examiner, and not shown to be cancer antigens. With respect to the failure of protein and mRNA levels to correlate for p53, again, this is irrelevant. Applicants have not claimed a quantitative assay. The correct question is: is the molecule expressed in cancer patients and usable for diagnosis? The correlation of p53 with cancer is too well known to comment on.

With respect to the examiner's hypothetical argument that "antibodies bind to epitopes in proteins and antibodies can cross react with different proteins containing the same or similar proteins," this again ignores the evidence. In the SEREX methodology, antibodies against a host cell, e.g., *E. coli*, are removed in a stripping step. All that is left are antibodies which bind to non-*E. coli* proteins. Applicants used a cDNA library. It is well known that, when using a cDNA library, a single cDNA molecule is used as the transfectant. Hence, if an antibody reacts with a transfected or transformed cell, it must be reacting with the foreign protein, since antibodies to native proteins have been eliminated. Hence, all of the examiner's points are based upon generalizations which do not stand up to analysis when the data are considered. With respect to the statement bridging pages 4-5:

LUD 5538.1 CIP - JEL/NDH (09807399)

"The data indicating that the KOC-2 polynucleotide was restricted to the normal testis and that the KOC-3 polynucleotide was universal in normal tissues."

This is absolutely an incorrect reading of the data. First of all, there were antibodies against KOC-2 and KOC-3 found in the patient (NW38) sample. Hence, there was no immunotolerance. Second, the data state that "the only positive normal tissue was testis." These do not state that this was the only positive tissue. There is a very big difference. Further, if the examiner is going to argue that levels of expression can vary, then she must concede that, notwithstanding expression of KOC-3 in normal tissue, cancer tissues express it at a level such that an immune response is possible - especially since applicants have shown this.

Turning next to the rejection under 35 USC § 112, the statement of this rejection is not consistent. All claims are rejected, even though the examiner states that the claims are enabling for "polynucleotides comprising SEQ ID NOS: 5-8..." If this is the case, then claims 54-70 and 108 and 109 should not be rejected, as claims 54-70 expressly recite what is said to be enabled, and claims 108 and 109 have nothing to do with the issues raised in the rejection. The rejection recites points "A" and "B," and it is noted that only claims 53, 71-73 and 80-84 are rejected. They are no reasons given for rejecting the remaining claims.

Applicants contacted the examiner regarding this issue on October 22. The undersigned was advised that, if the utility rejection were overcome, then the issue at point 7 would be moot. It is noted that claims 54-70 are not rejected in the discussion following point 7. Claims 53 & 54 were combined to form claim 110. As such, claim 110 and claims dependent thereon cannot be rejected under 35 USC §112.

In view of the foregoing, allowance of this application is believed proper and is urged.

LUD 5538.1 CIP - JEL/NDH (09807399)

In view of the lack of clarity in this rejection, applicants cannot address it. The examiner is invited to clarify the rejection so it can be addressed. At such time, applicants will respond fully.

Respectfully submitted,

FULBRIGHT & JAWORSKI, L.L.P.



Norman D. Hanson, Esq.

Registration No. 30,946

666 Fifth Avenue  
New York, New York 10103-3198  
Telephone: 212-318-3168  
Telecopier: 212-318-3400

<110> Chen, Yao-Tseng  
Gure, Ali  
Tsang, Solam  
Stockert, Elisabeth  
Jager, Elke  
Knuth, Alexander  
Old, Lloyd J.

<120> Isolated Nucleic Acid Molecules Encoding Cancer Associated Antigen, The  
Antigens Per Se, And Uses Thereof

<130> LUD 5538.1

<140> 09/270,437

<141> 1999 - 03 - 16

<150> 09/061,709

<151> 1998-04-17

<160> 23

<210> 1  
<211> 4265  
<212> DNA  
<213> Homo sapiens  
<220>  
<400> 1

GTCTGAAGGA CCTGAGGCAT TTTGTGACGA GGATCGTCTC AGGTCAGCGG AGGGAGGAGA 60  
CTTATAGACC TATCCAGTCT TCAAGGTGCT CCAGAAAGCA GGAGTTGAAG ACCTGGGTGT 120  
GAGGGACACA TACATCCTAA AAGCACCACA GCAGAGGAGG CCCAGGCAGT GCCAGGAGTC 180  
AAGGTTCCCA GAAGACAAAC CCCCTAGGAA GACAGGCCGAC CTGTGAGGCC CTAGAGCACC 240  
ACCTTAAGAG AAGAAGAGCT GTAAGCCGCG CTTTGTGAGA GCCATCATGG GGGACAAGGA 300  
TATGCCTACT GCTGGGATGC CGAGTCTTCT CCAGAGTTCC TCTGAGAGTC CTCAGAGTTG 360  
TCCTGAGGGG GAGGACTCCC AGTCTCCTCT CCAGATTCCC CAGAGTTCTC CTGAGAGCGA 420  
CGACACCCTG TATCCTCTCC AGAGTCCTCA GAGTCGTTCT GAGGGGGAGG ACTCCTCGGA 480  
TCCTCTCCAG AGACCTCCTG AGGGGAAGGA CTCCAGTCT CCTCTCCAGA TTCCCCAGAG 540  
TTCTCCTGAG GGCGACGACA CCCAGTCTCC TCTCCAGAAT TCTCAGAGTT CTCCTGAGGG 600  
GAAGGACTCC CTGTCTCCTC TAGAGATTTC TCAGAGCCCT CCTGAGGGTG AGGATGTCCA 660  
GTCTCCTCTG CAGAATCCTG CGAGTTCCTT CTCTCCTCT GCTTTATTGA GTATTTTCCA 720  
GAGTTCCTCT GAGAGTATTC AAAGTCCTTT TGAGGGTTTT CCCCAGTCTG TTCTCCAGAT 780  
TCCTGTGAGC GCCGCCTCCT CCTCCACTTT AGTGAGTATT TTCCAGAGTT CCCCTGAGAG 840  
TACTCAAAGT CCTTTTGAGG GTTTTCCCCA GTCTCCACTC CAGATTCCCTG TGAGCCGCTC 900  
CTTCTCCTCC ACTTTATTGA GTATTTTCCA GAGTTCCTCC GAGAGAAGTC AGAGAACTTC 960  
TGAGGGTTTT GCACAGTCTC CTCTCCAGAT TCCTGTGAGC TCCTCCTCGT CCTCCACTTT 1020  
ACTGAGTCTT TTCCAGAGTT CCCCTGAGAG AACTCAGAGT ACTTTTGAGG GTTTTCCCCA 1080  
GTCTCCACTC CAGATTCCCTG TGAGCCGCTC CTTCTCCTCC ACTTTATTGA GTATTTTCCA 1140  
GAGTTCCTCT GAGAGAATTC AGAGTACTTT TGAGGGTTTT GCCCAGTCTC CTCTCCAGAT 1200  
TCCTGTGAGC CCTCCTTCT CCTCCACTTT AGTGAGTATT TTCCAGAGTT CCCCTGAGAG 1260  
AACTCAGAGT ACTTTTGAGG GTTTTCCCCA GTCTCCTCTC CAGATTCCCTG TGAGCTCCTC 1320  
CTTCTCCTCC ACTTTATTGA GTCTTTTCCA GAGTTCCTCT GAGAGAATTC AGAGTACTTT 1380  
TGAGGGTTTT CCCCAGTCTC CTCTCCAGAT TCCTGGAAGC CCTCCTTCT CCTCCACTTT 1440  
ACTGAGTCTT TTCCAGAGTT CCCCTGAGAG AACTCACAGT ACTTTTGAGG GTTTTCCCCA 1500  
GTCTCCTCTC CAGATTCCCTA TGACCTCCTC CTCTCCTCT ACTTTATTGA GTATTTTACA 1560  
GAGTTCCTCT GAGAGTGCTC AAAGTGCTTT TGAGGGTTTT CCCCAGTCTC CTCTCCAGAT 1620  
TCCTGTGAGC TCCTCTTCT CCTACACTTT ATTGAGTCTT TTCCAGAGTT CCCCTGAGAG 1680  
AACTCACAGT ACTTTTGAGG GTTTTCCCCA GTCTCCTCTC CAGATTCCCTG TGAGCTCCTC 1740

CTCCTCCTCC TCCACTTTAT TGAGTCTTTT CCAGAGTTCC CCTGAGTGTA CTCAAAGTAC 1800  
TTTTGAGGGT TTTCCCCAGT CTCCTCTCCA GATTCCCTCAG AGTCCTCCTG AAGGGGAGAA 1860  
TACCCATTCT CCTCTCCAGA TTGTTCCAAG TCTTCCTGAG TGGGAGGACT CCCTGTCTCC 1920  
TCACTACTTT CCTCAGAGCC CTCCTCAGGG GGAGGACTCC CTATCTCCTC ACTACTTTCC 1980  
TCAGAGCCCT CCTCAGGGGG AGGACTCCCT GTCTCCTCAC TACTTTCTCTC AGAGCCCTCA 2040  
GGGGGAGGAC TCCCTGTCTC CTCACTACTT TCCTCAGAGC CCTCCTCAGG GGGAGGACTC 2100  
CATGTCTCCT CTCTACTTTT CTCAGAGTCC TCTTCAGGGG GAGGAATTCC AGTCTTCTCT 2160  
CCAGAGCCCT GTGAGCATCT GCTCCTCCTC CACTCCATCC AGTCTTCCCC AGAGTTTCCC 2220  
TGAGAGTTCT CAGAGTCCTC CTGAGGGGCC TGTCCAGTCT CCTCTCCATA GTCCTCAGAG 2280  
CCCTCCTGAG GGGATGCACT CCCAATCTCC TCTCCAGAGT CCTGAGAGTG CTCCTGAGGG 2340  
GGAGGATTCC CTGTCTCCTC TCCAAATTCC TCAGAGTCCT CTGAGGGGAG AGGACTCCCT 2400  
GTCTTCTCTC CATTTCCTC AGAGTCCCTC TGAGTGGGAG GACTCCCTCT TCTCTCCAGA GTCCTGTGAG 2460  
CTTCTCTCAG TTTCTCCTC AGGGGGAGGA CTTCAGTCT TCCCTCAGAG TCCCTGAGA GTCCTCAGAG 2520  
TATCTGCTCC TCCTCCACTT CTTTGAGTCT TCCCAGAGT TTCCTGAGA GTCCTCAGAG 2580  
TCCTCCTGAG GGGCCTGCTC AGTCTCCTCT CCAGAGACCT GTCAGTCTCT TCTTCTCTCA 2640  
CACTTTAGCG AGTCTTCTCC AAAGTTCCCA TGAGAGTCCCT CAGAGTCTCT CTGAGGGGCC 2700  
TGCCAGTCT CCTCTCCAGA GTCCTGTGAG CTCCTTCCCC TCCTCCACTT CATCGAGTCT 2760  
TTCCCAGAGT TCTCCTGTGA GTCCTTCCCC TCCTCCACTT TCATCGAGTC TTTCCAAGAG 2820  
TTCCCCTGAG AGTCTCTCTC AGAGTCTCTG GATCTCCTTC TCCTCCTCCA CTTCATTGAG 2880  
CCCATTCACT GAAGAGTCCA GCAGCCAGT AGATGAATAT ACAAGTTCCT CAGACACCTT 2940  
GCTAGAGAGT GATTCTCTGA CAGACAGCGA GTCCTTGATA GAGAGCGAGC CTTGTGTCAC 3000  
TTATACACTG GATGAAAAGG TGGACGAGTT GGCGCGGTTT CTTCTCCTCA AATATCAAGT 3060  
GAAGCAGCCT ATCACAAGG CAGAGATGCT GACGAATGTC ATCAGCAGGT ACACGGGCTA 3120  
CTTTCTCTG ATCTTCAGGA AAGCCCGTGA GTTCATAGAG ATACTTTTGT GCATTTCCT 3180  
GAGAGAAGTG GACCTGATG ACTCCTATGT CTTTGTAAC ACATTAGACC TCACCTCTGA 3240  
GGGGTGTCTG AGTGATGAGC AGGGCATGTC CCAGAACCGC CTCCTGATTC TTATTCTGAG 3300  
TATCATCTTC ATAAAGGGCA CCTATGCCTC TGAGGAGGTC ATCTGGGATG TGCTGAGTGG 3360  
AATAGGGGTG CGTGCTGGGA GGGAGCACTT TGCCTTTGGG GAGCCCAGGG AGCTCCTCAC 3420  
TAAAGTTTGG GTGCAGGAAC ATTACCTAGA GTACCGGGAG GTGCCCAACT CTTCTCCTCC 3480  
TCGTTACGAA TTCTGTGGG GTCCAAGAGC TCATTGAGAA GTCATTAAGA GGAAAGTAGT 3540  
AGAGTTTTTG GCCATGCTAA AGAATACCGT CCCTATTACC TTTCCATCCT CTTACAAGGA 3600  
TGCTTTGAAA GATGTGGAAG AGAGAGCCCA GGCCATAATT GACACCACAG ATGATTGAGC 3660  
TGCCACAGAA AGTGCAAGCT CCAGTGTCTAT GTCCCCCAGC TTCTCTTCTG AGTGAAGTCT 3720  
AGGGCAGATT CTTCCTCTG AGTTTGAAGG GGGCAGTCTGA GTTCTACGT GGTGAGGGC 3780  
CTGGTTGAGG CTGGAGAGAA CACAGTGCTA TTTGCATTTT TGTTCCATAT GGGTAGTTAT 3840  
GGGGTTTACC TGTTTTACTT TTGGGTATTT TTCAAATGCT TTTCTATTA ATAACAGGTT 3900  
TAAATAGCTT CAGAACTCTA GTTTATGCAC ATGAGTCGCA CATGTATTGC TGTTTTCTG 3960  
GTTTAAGAGT AACAGTTTGA TATTTTGTAA AAACAAAAC ACACCCAAAC ACACCACATT 4020  
GGGAAAACCT TCTGCCTCAT TTTGTGATGT GTCACAGGTT AATGTGGTGT TACTGTAGGA 4080  
ATTTCTTTGA AACTGTGAAG GAACTCTGCA GTTAAATAGT GGAATAAAGT AAAGGATTGT 4140  
TAATGTTTGC ATTTCTCTAG GTCCTTTAGT CTGTTGTTCT TGAAACTAA AGATACATAC 4200  
CTGGTTTGCT TGGCTTACGT AAGAAAGTCG AAGAAAGTAA ACTGTAATAA ATAAAAGTGT 4260  
CAGTG

<210> 2  
<211> 1142  
<212> PRT  
<213> Homo sapiens  
<220>  
<400> 2

Met Gly Asp Lys Asp Met Pro Thr Ala Gly Met Pro Ser Leu Leu Gln  
5 10 15  
Ser Ser Ser Glu Ser Pro Gln Ser Cys Pro Glu Gly Glu Asp Ser Gln  
20 25 30  
Ser Pro Leu Gln Ile Pro Gln Ser Pro Glu Ser Asp Asp Thr Leu  
35 40 45  
Tyr Pro Leu Gln Ser Pro Gln Ser Arg Ser Glu Gly Glu Asp Ser Ser  
50 55 60  
Asp Pro Leu Gln Arg Pro Pro Glu Gly Lys Asp Ser Gln Ser Pro Leu

65 70 75 80  
 Gln Ile Pro Gln Ser Ser Pro Glu Gly Asp Asp Thr Gln Ser Pro Leu  
 85 90 95  
 Gln Asn Ser Gln Ser Ser Pro Glu Gly Lys Asp Ser Leu Ser Pro Leu  
 100 105 110  
 Glu Ile Ser Gln Ser Pro Pro Glu Gly Glu Asp Val Gln Ser Pro Leu  
 115 120 125  
 Gln Asn Pro Ala Ser Ser Phe Phe Ser Ser Ala Leu Leu Ser Ile Phe  
 130 135 140  
 Gln Ser Ser Pro Glu Ser Ile Gln Ser Pro Phe Glu Gly Phe Pro Gln  
 145 150 155 160  
 Ser Val Leu Gln Ile Pro Val Ser Ala Ala Ser Ser Ser Thr Leu Val  
 165 170 175  
 Ser Ile Phe Gln Ser Ser Pro Glu Ser Thr Gln Ser Pro Phe Glu Gly  
 180 185 190  
 Phe Pro Gln Ser Pro Leu Gln Ile Pro Val Ser Arg Ser Phe Ser Ser  
 195 200 205  
 Thr Leu Leu Ser Ile Phe Gln Ser Ser Pro Glu Arg Ser Gln Arg Thr  
 210 215 220  
 Ser Glu Gly Phe Ala Gln Ser Pro Leu Gln Ile Pro Val Ser Ser Ser  
 225 230 235 240  
 Ser Ser Ser Thr Leu Leu Ser Leu Phe Gln Ser Ser Pro Glu Arg Thr  
 245 250 255  
 Gln Ser Thr Phe Glu Gly Phe Pro Gln Ser Pro Leu Gln Ile Pro Val  
 260 265 270  
 Ser Arg Ser Phe Ser Ser Thr Leu Leu Ser Ile Phe Gln Ser Ser Pro  
 275 280 285  
 Glu Arg Thr Gln Ser Thr Phe Glu Gly Phe Ala Gln Ser Pro Leu Gln  
 290 295 300  
 Ile Pro Val Ser Pro Ser Phe Ser Ser Thr Leu Val Ser Ile Phe Gln  
 305 310 315 320  
 Ser Ser Pro Glu Arg Thr Gln Ser Thr Phe Glu Gly Phe Pro Gln Ser  
 325 330 335  
 Pro Leu Gln Ile Pro Val Ser Ser Ser Phe Ser Ser Thr Leu Leu Ser  
 340 345 350  
 Leu Phe Gln Ser Ser Pro Glu Arg Thr Gln Ser Thr Phe Glu Gly Phe  
 355 360 365  
 Pro Gln Ser Pro Leu Gln Ile Pro Gly Ser Pro Ser Phe Ser Ser Thr  
 370 375 380  
 Leu Leu Ser Leu Phe Gln Ser Ser Pro Glu Arg Thr His Ser Thr Phe  
 385 390 395 400  
 Glu Gly Phe Pro Gln Ser Pro Leu Gln Ile Pro Met Thr Ser Ser Phe  
 405 410 415  
 Ser Ser Thr Leu Leu Ser Ile Leu Gln Ser Ser Pro Glu Ser Ala Gln  
 420 425 430  
 Ser Ala Phe Glu Gly Phe Pro Gln Ser Pro Leu Gln Ile Pro Val Ser  
 435 440 445  
 Ser Ser Phe Ser Tyr Thr Leu Leu Ser Leu Phe Gln Ser Ser Pro Glu  
 450 455 460  
 Arg Thr His Ser Thr Phe Glu Gly Phe Pro Gln Ser Pro Leu Gln Ile  
 465 470 475 480  
 Pro Val Ser Ser Ser Ser Ser Ser Thr Leu Leu Ser Leu Phe Gln  
 485 490 495  
 Ser Ser Pro Glu Cys Thr Gln Ser Thr Phe Glu Gly Phe Pro Gln Ser  
 500 505 510  
 Pro Leu Gln Ile Pro Gln Ser Pro Pro Glu Gly Glu Asn Thr His Ser  
 515 520 525  
 Pro Leu Gln Ile Val Pro Ser Leu Pro Glu Trp Glu Asp Ser Leu Ser  
 530 535 540



Pro His Tyr Phe Pro Gln Ser Pro Pro Gln Gly Glu Asp Ser Leu Ser  
 545 550 555 560  
 Pro His Tyr Phe Pro Gln Ser Pro Pro Gln Gly Glu Asp Ser Leu Ser  
 565 570 575  
 Pro His Tyr Phe Pro Gln Ser Pro Gln Gly Glu Asp Ser Leu Ser Pro  
 580 585 590  
 His Tyr Phe Pro Gln Ser Pro Pro Gln Gly Glu Asp Ser Met Ser Pro  
 595 600 605  
 Leu Tyr Phe Pro Gln Ser Pro Leu Gln Gly Glu Glu Phe Gln Ser Ser  
 610 615 620  
 Leu Gln Ser Pro Val Ser Ile Cys Ser Ser Ser Thr Pro Ser Ser Leu  
 625 630 635 640  
 Pro Gln Ser Phe Pro Glu Ser Ser Gln Ser Pro Pro Glu Gly Pro Val  
 645 650 655  
 Gln Ser Pro Leu His Ser Pro Gln Ser Pro Pro Glu Gly Met His Ser  
 660 665 670  
 Gln Ser Pro Leu Gln Ser Pro Glu Ser Ala Pro Glu Gly Glu Asp Ser  
 675 680 685  
 Leu Ser Pro Leu Gln Ile Pro Gln Ser Pro Leu Glu Gly Glu Asp Ser  
 690 695 700  
 Leu Ser Ser Leu His Phe Pro Gln Ser Pro Pro Glu Trp Glu Asp Ser  
 705 710 715 720  
 Leu Ser Pro Leu His Phe Pro Gln Phe Pro Pro Gln Gly Glu Asp Phe  
 725 730 735  
 Gln Ser Ser Leu Gln Ser Pro Val Ser Ile Cys Ser Ser Ser Thr Ser  
 740 745 750  
 Leu Ser Leu Pro Gln Ser Phe Pro Glu Ser Pro Gln Ser Pro Pro Glu  
 755 760 765  
 Gly Pro Ala Gln Ser Pro Leu Gln Arg Pro Val Ser Ser Phe Phe Ser  
 770 775 780  
 Tyr Thr Leu Ala Ser Leu Leu Gln Ser Ser His Glu Ser Pro Gln Ser  
 785 790 795 800  
 Pro Pro Glu Gly Pro Ala Gln Ser Pro Leu Gln Ser Pro Val Ser Ser  
 805 810 815  
 Phe Pro Ser Ser Thr Ser Ser Ser Leu Ser Gln Ser Ser Pro Val Ser  
 820 825 830  
 Ser Phe Pro Ser Ser Thr Ser Ser Ser Leu Ser Lys Ser Ser Pro Glu  
 835 840 845  
 Ser Pro Leu Gln Ser Pro Val Ile Ser Phe Ser Ser Ser Thr Ser Leu  
 850 855 860  
 Ser Pro Phe Ser Glu Glu Ser Ser Ser Pro Val Asp Glu Tyr Thr Ser  
 865 870 875 880  
 Ser Ser Asp Thr Leu Leu Glu Ser Asp Ser Leu Thr Asp Ser Glu Ser  
 885 890 895  
 Leu Ile Glu Ser Glu Pro Leu Phe Thr Tyr Thr Leu Asp Glu Lys Val  
 900 905 910  
 Asp Glu Leu Ala Arg Phe Leu Leu Lys Tyr Gln Val Lys Gln Pro  
 915 920 925  
 Ile Thr Lys Ala Glu Met Leu Thr Asn Val Ile Ser Arg Tyr Thr Gly  
 930 935 940  
 Tyr Phe Pro Val Ile Phe Arg Lys Ala Arg Glu Phe Ile Glu Ile Leu  
 945 950 955 960  
 Phe Gly Ile Ser Leu Arg Glu Val Asp Pro Asp Asp Ser Tyr Val Phe  
 965 970 975  
 Val Asn Thr Leu Asp Leu Thr Ser Glu Gly Cys Leu Ser Asp Glu Gln  
 980 985 990  
 Gly Met Ser Gln Asn Arg Leu Leu Ile Leu Ile Leu Ser Ile Ile Phe  
 995 1000 1005  
 Ile Lys Gly Thr Tyr Ala Ser Glu Glu Val Ile Trp Asp Val Leu Ser  
 1010 1015 1020

Gly Ile Gly Val Arg Ala Gly Arg Glu His Phe Ala Phe Gly Glu Pro  
 1025 1030 1035 1040  
 Arg Glu Leu Leu Thr Lys Val Trp Val Gln Glu His Tyr Leu Glu Tyr  
 1045 1050 1055  
 Arg Glu Val Pro Asn Ser Ser Pro Pro Arg Tyr Glu Phe Leu Trp Gly  
 1060 1065 1070  
 Pro Arg Ala His Ser Glu Val Ile Lys Arg Lys Val Val Glu Phe Leu  
 1075 1080 1085  
 Ala Met Leu Lys Asn Thr Val Pro Ile Thr Phe Pro Ser Ser Tyr Lys  
 1090 1095 1100  
 Asp Ala Leu Lys Asp Val Glu Glu Arg Ala Gln Ala Ile Ile Asp Thr  
 1105 1110 1115 1120  
 Thr Asp Asp Ser Thr Ala Thr Glu Ser Ala Ser Ser Ser Val Met Ser  
 1125 1130 1135  
 Pro Ser Phe Ser Ser Glu  
 1140

<210> 3  
 <211> 7  
 <212> PRT  
 <213> Homo sapiens  
 <220>  
 <400> 3

Pro Gln Ser Pro Leu Gln Ile  
 1 5

<210> 4  
 <211> 4159  
 <212> DNA  
 <213> Homo sapiens  
 <220>  
 <400> 4

GGTGGATGCG TTTGGGTTGT AGCTAGGCTT TTTCTTTTCT TTCTCTTTTA AAACACATCT 60  
 AGACAAGGAA AAAACAAGCC TCGGATCTGA TTTTTCACCTC CTCGTTCTTG TGCTTGGTTC 120  
 TTACTGTGTT TGTGTATTTT AAAGGCGAGA AGACGAGGGG AACAAAACCA GCTGGATCCA 180  
 TCCATCACCG TGGGTGGTTT TAATTTTTCG TTTTTCCTCG TTATTTTTCG TTAACAACCC 240  
 ACTCTTCACA ATGAACAAAC TGTATATCGG AAACCTCAGC GAGAACGCCG CCCCTCGGA 300  
 CCTAGAAAGT ATCTTCAAGG ACGCCAAGAT CCGGTGTGCG GGACCCTTCC TGGTGAAGAC 360  
 TGGCTACGCG TTCGTGGACT GCCCGGACGA GAGCTGGGCC CTCAAGGCCA TCGAGGCGCT 420  
 TTCAGGTAAT ATAGAACTGC ACGGGAAACC CATAGAAGTT GAGCACTCGG TCCCAAAAAG 480  
 GCAAAGGATT CGGAAACTTC AGATACGAAA TATCCCGCCT CATTTACAGT GGGAGGTGCT 540  
 GGATAGTTTA CTAGTCCAGT ATGGAGTGGT GGAGAGCTGT GAGCAAGTGA ACACTGACTC 600  
 GGAAACTGCA GTTGTAATG TAACCTATTC CAGTAAGGAC CAAGCTAGAC AAGCACTAGA 660  
 CAAACTGAAT GGATTTTCAGT TAGAGAATTT CACCTTGAAA GTAGCCTATA TCCCTGATGA 720  
 AATGGCCGCC CAGCAAAACC CTTGTCAGCA GCCCGGAGGT CGCCGGGGGC TTGGGCAGAG 780  
 GGGCTCCTCA AGGCAGGGGT CTCCAGGATC CGTATCCAAG CAGAAACCAT GTGATTTGCC 840  
 TCTGCGCCTG CTGGTTCCCA CCCAATTTGT TGGAGCCATC ATAGGAAAAG AAGGTGCCAC 900  
 CATTCGGAAC ATCACCACAC AGACCCAGTC TAAATTCGAT GTCCACCGTA AAGAAAATGC 960  
 GGGGGCTGCT GAGAAGTCGA TTACTATCCT CTCTACTCCT GAAGGACCT CTGCGGCTTG 1020  
 TAAGTCTATT CTGGAGATTA TGCATAAGGA AGCTAAGAT ATAAATTC AAGAAGGAAG 1140  
 CCCCTTGAAG ATTTTAGCTC ATAATAACTT TGTGAGCGT CTTATTGGTA AAGAAGGAAG 1200  
 AAATCTTAAA AAAATTGAGC AAGACACAGA CACTAAAATC ACGATATCTC CATTGCAGGA 1260  
 ATTGACGCTG TATAATCCAG AACGCACTAT TACAGTTAAA GGCAATGTTG AGACATGTGC 1320  
 CAAAGCTGAG GAGGAGATCA TGAAGAAAAT CAGGGAGTCT TATGAAATG ATATTGCTTC 1380  
 TATGAATCTT CAAGCACATT TAATTCCTGG ATTAATCTG AACGCCTTGG GTCTGTTCCC

ACCCACTTCA GGGATGCCAC CTCCACCTC AGGGCCCCCT TCAGCCATGA CTCCTCCCTA 1440  
 CCGGCAGTTT GAGCAATCAG AAACGGAGAC TGTTCATCAG TTTATCCCAG CTCTATCAGT 1500  
 CCGTGCCATC ATCGGCAAGC AGGGCCAGCA CATCAAGCAG CTTTCTCGCT TTGCTGGAGC 1560  
 TTCAATTAAG ATTGCTCCAG CGGAAGCACC AGATGCTAAA GTGAGGATGG TGATTATCAC 1620  
 TGGACCACCA GAGGCTCAGT TCAAGGCTCA GGAAGAATT TATGGAAAAA TTAAGAAGA 1680  
 AACTTTTGT AGTCCTAAAG AAGAGGTGAA ACTTGAAGCT CATATCAGAG TGCCATCCTT 1740  
 TGCTGCTGGC AGAGTTATTG GAAAAGGAGG CAAAACGGTG AATGAACCTC AGAATTTGTC 1800  
 AAGTGCAGAA GTTGTGTGCC CTCGTGACCA GACACCTGAT GAGAATGACC AAGTGGTTGT 1860  
 CAAAATAACT GGTCACTTCT ATGCTTGCCA GGTGCCCCAG AGAAAAATTC AGGAAATTC 1920  
 GACTCAGGTA AAGCAGCACC AACAACAGAA GGCTCTGCAA AGTGGACCAC CTCAGTCAAG 1980  
 ACGGAAGTAA AGGCTCAGGA AACAGCCCAC CACAGAGGCA GATGCCAAAC CAAAGACAGA 2040  
 TTGCTTAACC AACAGATGGG CGCTGACCCC CTATCCAGAA TCACATGCAC AAGTTTTTAC 2100  
 CTAGCCAGTT GTTTCTGAGG ACCAGGCAAC TTTTGAATC CTGTCTCTGT GAGAATGTAT 2160  
 ACTTTATGCT CTCTGAAATG TATGACACC AGCTTTAAAA CAAACAAACA AACAAACAAA 2220  
 AAAAGGGTGG GGGAGGGAGG GAAAGAGAAG AGCTCTGCAC TTCCCTTTGT TGTAAGTCTCA 2280  
 CAGTATAACA GATATTCTAA TTCTTCTTAA TATCCCCCA TAATGCCAGA AATTGGCTTA 2340  
 ATGATGCTTT CACTAAATTC ATCAAATAGA TTGCTCCTAA ATCCAATTGT TAAAATTGGA 2400  
 TCAGATAAT TATCAGAGGA ACTTAAATGT TAAGCCATTA GCATAGAAAA ACTGTTCTCA 2460  
 GTTTTATTTT TACCTAACAC TAACATGAGT AACCTAAGGG AAGTGTGAA TGGTGTGGC 2520  
 AGGGGTATTA AACGTGCATT TTTACTCAAC TACCTCAGGT ATTCAGTAAT ACAATGAAAA 2580  
 GCAAATTTGT TCCTTTTTTT TGAATTTTT ATATACTTTA TAATGATAGA AGTCCAACCG 2640  
 TTTTTTAAAA AATAAATTTA AAATTTAACA GCAATCAGCT AACAGGCAA TTAAGATTTT 2700  
 TACTTCTGGC TGGTGACAGT AAAGCTGGAA AATTAATTTT AGGGTTTTTT GAGGCTTTTG 2760  
 ACACAGTTAT TAGTTAAATC AAATGTTCAA AAATACGGAG CAGTGCCCTAG TATCTGGAGA 2820  
 GCAGCACTAC CATTTATTCT TTCATTATA GTTGGGAAG TTTTGTGACG TACTAACAAA 2880  
 GTGGTTCGAG GAGATTTTGG AACGGCTGGT TTAATGGCT TCAGGAGACT TCAGTTTTTT 2940  
 GTTTAGCTAC ATGATTGAAT GCATAATAAA TGCTTTGTGC TTCTGACTAT CAATACCTAA 3000  
 AGAAAGTGCA TCAGTGAAGA GATGCAAGAC TTTCACTGA CTGGCAAAA GCAAGCTTTA 3060  
 GCTTGTCTTA TAGGATGCTT AGTTTGCCAC TACACTTCAG ACCAATGGGA CAGTCATAGA 3120  
 TGGTGTGACA GTGTTTAAAC GCAACAAAAG GCTACATTTT CATGGGGCCA GCCTGTCTAT 3180  
 GAGCCTCACT AAGCTATTTT GAAGATTTT AAGCACTGAT AAATTA AAAA 3240  
 AAATTAGACT CCACCTTAAG TAGTAAAGTA TAACAGGATT TCTGTATACT GTGCAATCAG 3300  
 TTCTTTGAAA AAAAAGTCAA AAGATAGAGA ATACAAGAAA AGTTTNGGG ATATAATTG 3360  
 AATGACTGTG AAAACATATG ACCTTTGATA ACGAACTCAT TTGCTCACTC CTTGACAGCA 3420  
 AAGCCAGTA CGTACAATTG TGTTGGGTGT GGGTGGTCTC CAAGGCCACG CTGCTCTCTG 3480  
 AATTGATTTT TTGAGTTTTG GNTTGNAAGA TGATCACAGN CATGTTACAC TGATCTTNA 3540  
 GGACATATNT TATAACCTTT TAAAAAAA ATCCCTGCC TCATTCTTAT TTCGAGATGA 3600  
 ATTTGATATC AGACTAGATG TCTTCTGAA GATCAATTAG ACATTNTGAA AATGATTTAA 3660  
 AGTGTTCCT TTAATGTTCT CTGAAAACAA GTTCTTTTG TAGTTTAAAC CAAAAAAGTG 3720  
 CCCTTTTGT CACTGGTTTC TCCTAGCATT CATGATTTT TTTTCAACA ATGAATTAA 3780  
 ATTGCTAAAA TCATGGACTG GCTTCTGTT TGGATTTTCA GTAGATGTG TTTAAGGCCA 3840  
 GAGCTTTTCT CAGTATTTGA TTTTTTCCC CAATATTTGA TTTTTTAAAA ATATACACAT 3900  
 AGGAGCTGCA TTTAAACCT GCTGGTTTAA ATTCTGTCA ATTTCACTTC TAGCCTTTTA 3960  
 GTATGGCNAA TCANAATTTA CTTTACTTA AGCATTGTG ATTTGGAGTA TCTGGTACTA 4020  
 GCTAAGAAAT AATTCNATAA TTGAGTTTG TACTCNCCAA ANATGGGTCA TTCCTCATGN 4080  
 ATAATGTNCC CCAATGCAG CTTCATTTTC CAGANACCTT GACGCAGGAT AAATTTTTTC 4140  
 ATCATTTAGG TCCCCAAA 4159

<210> 5  
 <211> 1708  
 <212> DNA  
 <213> Homo sapiens  
 <220>  
 <400> 5

AGGGACGCTG CCGCACCGCC CCAGTTTACC CCGGGGAGCC ATCATGAAGC TGAATGGCCA 60  
 CCAGTTGGAG AACCATGCC TGAAGGTCTC CTACATCCCC GATGAGCAGA TAGCACAGGG 120  
 ACCTGAGAAT GGGCGCCGAG GGGGCTTTGG CTCTCGGGGT CAGCCCCGCC AGGGCTCACC 180  
 TGTGCCAGCG GGGGCCCCAG CCAAGCAGCA GCAAGTGGAC ATCCCCCTTC GGCTCCTGGT 240

GCCCACCCAG TATGTGGGTG CCATTATTGG CAAGGAGGGG GCCACCATCC GCAACATCAC 300  
 AAAACAGACC CAGTCCAAGA TAGACGTGCA TAGGAAGGAG AACGCAGGTG CAGCTGAAAA 360  
 AGCCATCAGT GTGCACTCCA CCCCTGAGGG CTGCTCCTCC GCTTGTAAGA TGATCTTGGA 420  
 GATTATGCAT AAAGAGGCTA AGGACACCAA AACGGCTGAC GAGGTTCCCC TGAAGATCCT 480  
 GGCCCATAAAT AACTTTGTAG GCGTCTCAT TGGCAAGGAA GGACGGAACC TGAAGAAGGT 540  
 AGAGCAAGAT ACCGAGACAA AAATCACCAT CTCCTCGTTG CAAGACCTTA CCCTTTACAA 600  
 CCCTGAGAGG ACCATCACTG TGAAGGGGGC CATCGAGAAT TGTTCAGGG CCGAGCAGGA 660  
 AATAATGAAG AAAGTTCGGG AGGCCTATGA GAATGATGTG GCTGCCATGA GCTCTCACCT 720  
 GATCCCTGGC CTGAACCTGG CTGCTGTAGG TCTTTTCCCA GCTTCATCCA GCGCAGTCCC 780  
 GCCGCCTCCC AGCAGCGTTA CTGGGGCTGC TCCCTATAGC TCCTTTATGC AGGCTCCCGA 840  
 GCAGGAGATG GTGCAGGTGT TTATCCCCGC CCAGGCAGTG GGCGCCATCA TCGGCAAGAA 900  
 GGGGCAGCAC ATCAAACAGC TCTCCCGGTT TGCCAGCGCC TCCATCAAGA TTGCACCACC 960  
 CGAAACACCT GACTCCAAAG TTCGTATGGT TATCATCACT GGACCGCCAG AGGCCCAATT 1020  
 CAAGGCTCAG GGAAGAATCT ATGGCAAACT CAAGGAGGAG AACTTCTTTG GTCCCAAGGA 1080  
 GGAAGTGAAG CTGGAGACCC ACATACGTGT GCCAGCATCA GCAGCTGGCC GGGTCATTGG 1140  
 CAAAGGTGGA AAAACGGTGA ACGAGTTGCA GAATTTGACG GCAGCTGAGG TGGTAGTACC 1200  
 AAGAGACCAG ACCCCTGATG AGAACGACCA GGTCTCTGTG AAAATCATCG GACATTTCTA 1260  
 TGCCAGTCAG ATGGCTCAAC GGAAGATCCG AGACATCCTG GCCCAGGTTA AGCAGCAGCA 1320  
 TCAGAAGGGA CAGAGTAACC AGGCCAGGC ACGGAGGAAG TGACCAGCCC CTCCCTGTCC 1380  
 CTTNGAGTCC AGGACAACAA CACNTGGGCC GGGCTGTAGA TCAGGTTTGC CCACTTGATT 1500  
 GAATGAGTGG GAATCCGGGA CACNTGGGCC GGGCTGTAGA TCAGGTTTGC CCACTTGATT 1500  
 SAGAAAGATG TTCCAGTGAG GAACCTGAT CTNTCAGCCC CAAACACCCA CCCAATTGGC 1560  
 CCPACACTGT NTGCCCTCG GGGTGTGAGA AATTNTAGCG CAAGGCACTT TTAACCTGG 1620  
 ATTGTTTAAA GAAGCTCTCC AGGCCCCACC AAGAGGGTGG ATCACACCTC AGTGGGAAGA 1680  
 AAAATAAAAT TTCCTTCAGG TTTTAAAA 1708

<210> 6  
 <211> 3412  
 <212> DNA  
 <213> Homo sapiens  
 <220>  
 <400> 6

GGCAGCGGAG GAGGCGAGGA GCGCCGGGTA CCGGGCCGGG GGAGCCGCGG GCTCTCGGGG 60  
 AAGAGACCGA TGATGAACAA GCTTTACATC GGAACCTGA GCCCGCCGT CACCGCCGAC 120  
 GACCTCCGGC AGCTCTTTGG GGACAGGAAG CTGCCCTGG CGGGACAGGT CCTGCTGAAG 180  
 TCCGGCTACG CTTTCGTGGA CTACCCCGAC CAGAAGTGGG CCATCCGCGC CATCGAGACC 240  
 CTCTCGGGTA AAGTGAATTT GCATGGGAAA ATCATGGAAG TTGATTACTC AGTCTCTAAA 300  
 AAGCTAAGGA GCAGGAAAAAT TCAGATTGCA AACATCCCTC CTCACCTGCA GTGGGAGGTG 360  
 TTGGATGGAC TTTTGGCTCA ATATGGGACA GTGGAGAATG TGGAAACAGT CAACACAGAC 420  
 ACAGAAACCG CCGTTGTCAA CGTCACATAT GCAACAAGAG AAGAAGCAAA AATAGCCATG 480  
 GAGAAGCTAA GCGGGCATCA GTTTGAGAAC TACTCCTTCA AGATTTCCTA CATCCCGGAT 540  
 GAAAGAGTGA GCTCCCTTTC GCCCCTCAG CGAGCCAGC GTGGGGACCA CTCTTCCGG 600  
 GAGCAAGGCC ACGCCCTGG GGGCACTTCT CAGGCCAGAC AGATTGATTT CCCGCTGCGG 660  
 ATCCTGGTCC CCACCCAGTT TGTGGTGGC ATCATCGGAA AGGAGGGCTT GACCATAAAG 720  
 AACATCACTA AGCAGACCCA GTCCCGGGTA GATATCCATA GAAAAGAGAA CTCTGGAGCT 780  
 GCAGAGAAGC CTGTCAACCAT CCATGCCACC CCAGAGGGGA CTTCTGAAGC ATGCCCGCATG 840  
 ATTCTTGAAA TCATGCAGAA AGAGGCAGAT GAGACCAAAC TAGCCGAAGA GATTCTCTG 900  
 AAAATCTTGG CACACAATGG CTTGGTTGGA AGACTGATTG GAAAAGAAGG CAGAAATTG 960  
 AAGAAAATTG AACATGAAC AGGGACCAAG ATAACAATCT CATCTTTGCA GGATTGAGC 1020  
 ATATACAAAC CGGAAAGAAC CATCACTGTG AAGGGCACAG TTGAGGCTG TGCCAGTGCT 1080  
 GAGATAGAGA TTATGAAGAA GCTGCGTGAG GCCTTTGAAA ATGATATGCT GGCTGTAAAC 1140  
 CAACAAGCCA ATCTGATCCC AGGGTTGAAC CTCAGCGCAC TTGGCATCTT TTCAACAGGA 1200  
 CTGTCCGTGC TATCTCCACC AGCAGGGCCC CGCGGAGCTC CCCCCGCTGC CCCCTACCAC 1260  
 CCCTTCACTA CCCACTCCGG ATACTTCTCC AGCCTGTACC CCCATCACCA GTTTGGCCCCG 1320  
 TTCCCGCATC ATCACTCTTA TCCAGAGCAG GAGATTGTA ATCTCTCAT CCCAACCCAG 1380  
 GCTGTGGGCG CCATCATCGG GAAGAAGGGG GCACACATCA AACAGCTGGC GAGATTGCGC 1440  
 GGAGCCTCTA TCAAGATTGC CCCTGCGGAA GGCCAGAGC TCAGCGAAAG GATGGTCATC 1500  
 ATCACCGGGC CACCGGAAGC CCAGTCAAG GCCCAGGGAC GGATCTTTGG GAAACTGAAA 1560

```

GAGGAAAAC TCTTTAACCC CAAAGAAGAA GTGAAGCTGG AAGCGCATAT CAGAGTGCCC 1620
TCTTCCACAG CTGGCCGGGT GATTGGCAAA GGTGGCAAGA CCGTGAACGA ACTGCAGAAC 1680
TTAACCAGTG CAGAAGTCAT CGTGCCTCGT GACCAAACGC CAGATGAAAA TGAGGAAGTG 1740
ATCGTCAGAA TTATCGGGCA CTTCTTTGCT AGCCAGACTG CACAGCGCAA GATCAGGGAA 1800
ATTGTACAAC AGGTGAAGCA GCAGGAGCAG AAATACCCTC AGGGAGTCGC CTCACAGCGC 1860
AGCAAGTGAG GCTCCACAG GCACCAGCAA AACACCGAT GAATGTAGCC CTCCAACAC 1920
CTGACAGAAT GAGACCAAA GCAGCCAGCC AGATCGGGAG CAAACCAAAG ACCATCTGAG 1980
GAATGAGAAG TCTGCGGAG CGGCCAGGGA CTCTGCCGAG GCCCTGAGAA CCCAGGGGGC 2040
CGAGGAGGGG CGGGGAAGGT CAGCCAGGTT TGCCAGAACC ACCGAGCCCC GCCTCCCGCC 2100
CCCCAGGGCT TCTGCAGGCT TCAGCCATCC ACTTCACCAT CCACTCGGAT CTCTCCTGAA 2160
CTCCACAGAC GCTATCCCTT TTAGTTGAAC TAACATAGGT GAACGTGTTT AAAGCCAAGC 2220
AAAATGCACA CCCTTTTTCT GTGGCAAATC GTCTCTGTAC ATGTGTGTAC ATATTAGAAA 2280
GGGAAGATGT TAAGATATGT GGCCTGTGGG TTACACAGGG TGCCTGCAGC GGTAATATAT 2340
TTTTAGAAATA ATATATCAAA TAACTCAACT AACTCCAATT TTTAATCAAT TATTAATTTT 2400
TTTTCTTTT TAAAGAGAAA GCAGGCTTTT CTAGACTTTA AAGAATAAAG TCTTTGGGAG 2460
GTCTCAGGGT GTAGAGAGGA GCTTTGAGGC CACCCGCACA AAATTCACCC AGAGGGAAAT 2520
CTCGTCGGAA GGACACTCAC GGCAGTCTG GATCACCTGT GTATGTCAAC AGAAGGGATA 2580
CCGTCTCCTT GAAGAGGAAA CTCTGTCACT CCTCATGCCT GTCTAGCTCA TACACCCATT 2640
TCTCTTTGCT TCACAGGTTT TAAACTGGTT TTTTGCATAC TGCTATATAA TTCTCTGTCT 2700
CTCTCTGTTT ATCTCTCCCC TCCCTCCCTT CCCCTTCTTC TCCATCTCCA TTCTTTTGAA 2760
TTTCTCATC CCTCCATCTC AATCCCGTAT CTACGCACCC CCCCCCCCC AGGCAAGCA 2820
GTGCTCTGAG TATCACATCA CACAAAAGGA ACAAAGCGA AACACACAA CCAGCCTCA 2880
CTTACACTTG GTTACTCAA AGAACAAGAG TCAATGGTAC TTGTCCTAGC GTTTTGGAAG 2940
AGGAAAACAG GAACCCACCA AACCAACCAA TCAACCAAAC AAAGAAAAA TTCCACAATG 3000
AAAGAATGTA TTTTGTCTTT TTGCATTTTG GTGTATAAGC CATCAATATT CAGCAAAATG 3060
ATTCCTTTCT TTAATAAAAA AAATGTGGAG GAAAGTAGAA ATTTACCAAG GTTGTGCGC 3120
CAGGGCGTTA AATTCACAGA TTTTTTAAAC GAGAAAAACA CACAGAAGAA GCTACCTCAG 3180
GTGTTTTTAC CTCAGCACCT TGCTCTGTG TTTCCCTTAG AGATTTTGT AAGCTGATAG 3240
TTGGAGCATT TTTTATTTT TTTAATAAAA ATGAGTTGGA AAAAAAATA GATATCAACT 3300
GCCAGCCTGG AGAAGGTGAC AGTCCAAAGT TGCAACAGCT GTTCTGAATT GTCTTCCGCT 3360
AGCCAAGAAC CNATATGGCC TTCTTTTGGA CAAACCTTGA AAATGTTTAT TT 3412

```

<210> 7  
 <211> 1946  
 <212> DNA  
 <213> Homo sapiens  
 <220>  
 <400> 7

C2

```

GCTGTAGCGG AGGGGCTGGG GGGCTGCTCT GTCCCTTCC TTGCGCGCTG CGGCCTCAGC 60
CCACCCAGAG GCCGGGGTGG GAGGGCGAGT GCTCAGCTTC CCGGGTTAGG AGCCGGAAAA 120
TTCAAATCCG AAATATTCCA CCCCAGCTCC GATGGGAAGT ACTGGACAGC CTGCTGGCTC 180
AGTATGGTAC AGTAGAGAAC TGTGAGCAAG TGAACACCGA GAGTGAGACG GCAGTGGTGA 240
ATGTCACCTA TTCCAACCGG GAGCAGACCA GGCAAGCCAT CATGAAGCTG AATGGCCACC 300
AGTTGGAGAA CCATGCCCTG AAGGTCTCCT ACATCCCGA TGAGCAGATA GCACAGGGAC 360
CTGAGAATGG GCGCCGAGGG GGCTTTGGCT CTCGGGGTCA GCGCCGCCAG GGCTCACCTG 420
TGCGAGCGGG GGCCCAGCC AAGCAGCAGC AAGTGGACAT CCCCCTTCGG CTCCTGGTGC 480
CCACCCAGTA TGTGGGTGCC ATTATTGGCA AGGAGGGGGC CACCATCCGC AACATCACAA 540
AACAGACCCA GTCCAAGATA GACGTGCATA GGAAGGAGAA CGCAGGTGCA GCTGAAAAAG 600
CCATCAGTGT GCACTCCACC CCTGAGGGCT GCTCCTCCGC TTGTAAGATG ATCTTGGAGA 660
TTATGCATAA AGAGGCTAAG GACACCAAAA CGGCTGACGA GGTTCCTCTG AAGATCCTGG 720
CCCATAATAA CTTTGTAGGG CGTCTCATTG GCAAGGAAGG ACGGAACCTG AAGAAGGTAG 780
AGCAAGATAC CGAGACAAA ATCACCATCT CCTCGTTGCA AGACCTTACC CTTTACAACC 840
CTGAGAGGAC CATCACTGTG AAGGGGGCCA TCGAGAATTG TGCAGGGGCC GAGCAGGAAA 900
TAATGAAGAA AGTTCGGGAG GCCTATGAGA ATGATGTGGC TGCCATGAGC TCTCACCTGA 960
TCCCTGGCCT GAACCTGGCT GCTGTAGGTC TTTTCCAGC TTCATCCAGC GCAGTCCCGC 1020
CGCCTCCAG CAGCGTTACT GGGGCTGCTC CCTATAGCTC CTTTATGCAG GCTCCCGAGC 1080
AGGAGATGGT GCAGGTGTTT ATCCCCGCC AGGCAGTGGG CGCCATCATC GGCAAGAAGG 1140
GGCAGCACAT CAAACAGCTC TCCGGGTTTG CCAGCGCCTC CATCAAGATT GCACCACCCG 1200

```

AAACACCTGA CTCCAAAGTT CGTATGGTTA TCATCACTGG ACCGCCAGAG GCCCAATTCA 1260  
 AGGCTCAGGG AAGAATCTAT GGCAAACTCA AGGAGGAGAA CTTCTTTGGT CCCAAGGAGG 1320  
 AAGTGAAGCT GGAGACCCAC ATACGTGTGC CAGCATCAGC AGCTGGCCGG GTCATTGGCA 1380  
 AAGGTGGAAG AACGGTGAAC GAGTTGCAGA ATTTGACGGC AGCTGAGGTG GTAGTACCAA 1440  
 GAGACCAGAC CCCTGATGAG AACGACCAGG TCATCGTGAA AATCATCGGA CATTCTATG 1500  
 CCAGTCAGAT GGCTCAACGG AAGATCCGAG ACATCCTGGC CCAGGTAAAG CAGCAGCATC 1560  
 AGAAGGGACA GAGTAACCAG GCCCAGGCAC GGAGGAAGTG ACCAGCCCCT CCCTGTCCCT 1620  
 TNGAGTCCAG GACAACAACG GGCAGAAATC GAGAGTGTGC TCTCCCCGGC AGGCCTGAGA 1680  
 ATGAGTGGGA ATCCGGGACA CNTGGGCCGG GCTGTAGATC AGGTTTGGCC ACTTGATTGA 1740  
 GAAAGATGTT CCAGTGAGGA ACCCTGATCT NTCAGCCCCA AACACCCACC CAATTGGCCC 1800  
 AACACTGTNT GCCCCTCGGG GTGTCAAGAA TTNTAGCGCA AGGCACTTTT AAACGTGGAT 1860  
 TGTTTAAAGA AGCTCTCCAG GCCCCACCAA GAGGGTGGAT CACACCTCAG TGGGAAGAAA 1920  
 AATAAAATTT CTTTCAGGTT TAAAAA 1946

<210> 8  
 <211> 3283  
 <212> DNA  
 <213> Homo sapiens  
 <220>  
 <400> 8

GGCAGCGGAG GAGGCGAGGA GCGCCGGGTA CCGGGCCGGG GGAGCCGGG GCTCTCGGGG 60  
 AAGAGACGGA TGATGAACAA GCTTTACATC GGGAACTGA GCCCCGCCGT CACCGCCGAC 120  
 GACCTCCGGC AGCTCTTTGG GGACAGGAAG CTGCCCCTGG CGGGACAGGT CCTGCTGAAG 180  
 TCCGGCTACG COTTCTGTGA CTACCCCGAC CAGAACTGGG CCATCCGCGC CATCGAGACC 240  
 CTCTCGGGTA AAGTGGAAAT GCATGGGAAA ATCATGGAAG TTGATTACTC AGTCTCTAAA 300  
 AAGCTAAGGA GCAGGAAAAT TCAGATTCTGA AACATCCCTC CTCACCTGCA GTGGGAGGTG 360  
 TTGGATGGAC TTTTGGCTCA ATATGGGACA GTGGAGAATG TGGAAACAAGT CAACACAGAC 420  
 ACAGAAACCG CCGTTGTCAA CGTCACATAT GCAACAAGAG AAGAAGCAAA AATAGCCATG 480  
 GAGAAGCTAA GCGGGCATCA GTTTGAGAAT TACTCCTTCA AGATTTCTTA CATCCCGGAT 540  
 GAAGAGGTGA GCTCCCTTTC GCCCCTCAG CGAGCCAGC GTGGGGACCA CTCTTCCCGG 600  
 GAGCAAGGCC ACGCCCTGG GGGCACTTCT CAGGCCAGAC AGATTGATTT CCCGCTGCGG 660  
 ATCCTGGTCC CCAACCCAGTT TGTGGTGCC ATCATCGGAA AGGAGGGCTT GACCATAAAG 720  
 AACATCACTA AGCAGACCCA GTCCCGGTA GATATCCATA GAAAAGAGAA CTCTGGAGCT 780  
 GCAGAGAAGC CTGTCAACAT CCATGCCACC CCAGAGGGGA CTTCTGAAGC ATGCCGATG 840  
 ATTCTTGAAA TCATGCAGAA AGAGGCAGAT GAGACCAAA TAGCCGAAGA GATTCTCTG 900  
 AAAATCTTGG CACACAATGG CTTGGTTGGA AGACTGATTG GAAAAGAAGG CAGAAATTTG 960  
 AAGAAAATTG AACATGAAAC AGGGACCAAG ATAACAATCT CATCTTTGCA GGATTTGAGC 1020  
 ATATACAACC CGGAAAGAAC CATCACTGTG AAGGGCAGG TTGAGGCTG TGCCAGTGCT 1080  
 GAGATAGAGA TTATGAAGAA GCTGCGTGAG GCCTTTGAAA ATGATATGCT GGCTGTAAAC 1140  
 ACCCACTCCG GATACTTCTC CAGCCTGTAC CCCCATCACC AGTTTGGCCC GTTCCCGCAT 1200  
 CATCACTCTT ATCCAGAGCA GGAGATTGTG AATCTCTTCA TCCCAACCCA GGCTGTGGGC 1260  
 GCCATCATCG GGAAGAAGGG GGCACACATC AAACAGCTGG CGAGATTGCG CGGAGCCTCT 1320  
 ATCAAGATTG CCCCTGCGGA AGGCCAGAG GTCAGCGAAA GGATGGTCAT CATCACCAGG 1380  
 CCACCGGAAG CCCAGTTCAA GGCCCAGGA CGGATCTTTG GGAAACTGAA AGAGGAAAAC 1440  
 TTCTTTAACC CCAAAGAAGA AGTGAAGCTG GAAGCGCATA TCAGAGTGCC CTCTTCCACA 1500  
 GCTGGCCGGG TGATTGGCAA AGGTGGCAAG ACCGTGAACG AACTGCAGAA CTTAACCAAGT 1560  
 GCAGAAGTCA TCGTGCCCTG TGACCAACCG CCAGATGAAA ATGAGGAAGT GATCGTCAGA 1620  
 ATTATCGGGC ACTTCTTTGC TAGCCAGACT GCACAGCGCA AGATCAGGGA AATTGTACAA 1680  
 CAGGTGAAGC AGCAGGAGCA GAAATACCCT CAGGGAGTCG CCTCACAGCG CAGCAAGTGA 1740  
 GGCTCCACAC GGCACCAAGCA AAACAACGGA TGAATGTAGC CCTTCCAACA CCTGACAGAA 1800  
 TGAGACCAAA CGCAGCCAGC CAGATCGGGA GCAAACCAAA GACCATCTGA GGAATGAGAA 1860  
 GTCTGCGGAG GCGGCCAGGG ACTCTGCCGA GGCCCTGAGA ACCCCAGGGG CCGAGGAGGG 1920  
 GCGGGGAAGG TCAGCCAGGT TTGCCAGAAC CACCGAGCCC CGCCTCCCGC CCCCAGGGC 1980  
 TTCTGCAGGC TTCAGCCATC CACTTCACCA TCCACTCGGA TCTCTCTGA ACTCCACGA 2040  
 CGCTATCCCT TTTAGTTGAA CTAACATAGG TGAACGTGTT CAAAGCCAAG CAAAATGCAC 2100  
 ACCCTTTTTC TGTGGCAAAT CGTCTCTGTA CATGTGTGTA CATATTAGAA AGGGAAGATG 2160  
 TTAAGATATG TGGCCTGTGG GTTACACAGG GTGCCTGCAG CGGTAATATA TTTTAGAAAT 2220  
 AATATATCAA ATAACCAAC TAACTCCAAT TTTAATCAA TTATTAATTT TTTTCTCTT 2280

```

TTAAAGAGAA AGCAGGCTTT TCTAGACTTT AAAGAATAAA GTCTTTGGGA GGTCTCACGG 2340
TGTAGAGAGG AGCTTTGAGG CCACCCGCAC AAAATTACAC CAGAGGGGAA TCTCGTCGGA 2400
AGGACACTCA CGGCAGTTCT GGATCACCTG TGTATGTCAA CAGAAGGGAT ACCGTCTCCT 2460
TGAAGAGGAA ACTCTGTCAC TCCTCATGCC TGTCTAGCTC ATACACCCAT TTCTCTTTGC 2520
TTCACAGGTT TTAAACTGGT TTTTTCGATA CTGCTATATA ATTCTCTGTC TCTCTCTGTT 2580
TATCTCTCCC CTCCTCCCC TCCCCTCTT CTCCATCTCC ATTCTTTTGA ATTTCTCAT 2640
CCCTCCATCT CAATCCCGTA TCTACGCACC CCCCCCCCCC CAGGCAAAGC AGTGCTCTGA 2700
GTATCACATC ACACAAAAGG AACAAAAGCG AAACACACAA ACCAGCCTCA ACTTACACTT 2760
GGTTACTCAA AAGAACAAGA GTCAATGGTA CTTGTCCTAG CGTTTTGGAA GAGGAAAACA 2820
GGAACCCACC AAACCAACCA ATCAACCAA CAAAGAAAA ATTCCACAAT GAAAGAATGT 2880
ATTTTGTCTT TTGTCATTTT GGTGTATAAG CCATCAATAT TCAGCAAAT GATTCCTTTC 2940
TTTAAAAAAA AAAATGTGGA GGAAAGTAGA AATTTACCAA GGTGTGTTGGC CCAGGGCGTT 3000
AAATTACAGC ATTTTTTTAA CGAGAAAAC ACACAGAAGA AGCTACCTCA GGTGTTTTTA 3060
CCTCAGCACC TTGCTCTTGT GTTCCCTTA GAGATTTTGT AAAGCTGATA GTTGGAGCAT 3120
TTTTTTATTT TTTAATAAAA AATGAGTTGG AAAAAAATA AGATATCAAC TGCCAGCCTG 3180
GAGAAGGTGA CAGTCCAAGT GTGCAACAGC TGTTCTGAAT TGTCTCCGC TAGCCAAGAA 3240
CCNATATGGC CTTCTTTTGG ACAAACCTTG AAAATGTTTA TTT 3283

```

<210> 9  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
 <220>

<400> 9  
 gaagtatct tcaaggacgc c 21

<210> 10  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens  
 <220>

<400> 10  
 ctgcaagggg ttttgctggg cg 22

<210> 11  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens  
 <220>

<400> 11  
 tccttgcgog ctgcggcctc ag 22

<210> 12  
 <211> 23  
 <212> DNA  
 <213> Homo sapiens  
 <220>

<400> 12  
 ccaactggtg gccattcagc ttc 23

<210> 13  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens  
 <220>

<400> 13  
 gctctttggg gacaggaagg tc 22

<210> 14  
 <211> 22

<212> DNA  
 <213> Homo sapiens  
 <220>  
 <400> 14  
 gacgttgaca acggcggttt ct 22

<210> 15  
 <211> 12  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <221> adaptor  
 <222> 1...12  
 <400> 15  
 aatttgcggt ga 12

<210> 16  
 <211> 12  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <221> adaptor  
 <222> 1...12  
 <400> 16  
 aatttgttca tg 12

<210> 17  
 <211> 12  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <221> adaptor  
 <222> 1...12  
 <400> 17  
 aattttccct cg 12

C2  
 Cuf  
 <210> 18  
 <211> 24  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <221> adaptor  
 <222> 1...24  
 <400> 18  
 agcactctcc agcctctcac catg 24

<210> 19  
 <211> 24  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <221> adaptor  
 <222> 1...24  
 <400> 19  
 accgacgtcg actatcatg catg 24  
 <210> 20  
 <211> 24



<212> DNA  
<213> artificial sequence  
<220>  
<221> adaptor  
<222> 1...24  
<400> 20

aggcaactgt gctatccgag catg 24

<210> 21  
<211> 8  
<212> DNA  
<213> artificial sequence  
<220>  
<221> adaptor  
<222> 1...8  
<400> 21

gtgagagg 8

<210> 22  
<211> 8  
<212> DNA  
<213> artificial sequence  
<220>  
<221> adaptor  
<222> 1...8  
<400> 22

catggatg 8

C2  
✓  
<210> 23  
<211> 8  
<212> DNA  
<213> artificial sequence  
<220>  
<221> adaptor  
<222> 1...8  
<400> 23

ctcggata 8